**Draft HUMAnN3 Protocol**

Protocol by Modupe Coker & Yuka Moroishi

Code/Set Up by Kevin Bonham, Modupe Coker, Yuka Moroishi, & Scot Zens

**Installation:**

Running HUMAnN3 requires Python3. Version used for first iteration is Python 3.7.8.

Refer to HUMAnN3 tutorial for specific installation instructions. Link to HUMAnN3 tutorial: https://github.com/biobakery/biobakery/wiki/humann3

Packages to install:

1. MetaPhlAn: https://huttenhower.sph.harvard.edu/metaphlan3/
2. HUMAnN: HUMAnN3 tutorial: https://github.cxom/biobakery/biobakery/wiki/humann3
3. Snakemake: https://snakemake.readthedocs.io/en/stable/getting\_started/installation.html

Example code:

$ conda create --name biobakery3 python=3.7

$ conda activate biobakery3

$ conda config --add channels defaults

$ conda config --add channels bioconda

$ conda config --add channels conda-forge

$ conda config --add channels biobakery

$ conda install humann -c biobakery

Test installation:

$ humann\_test

Databases are downloaded in the following locations in ‘/dartfs-hpc/rc/lab/M/MRKepistor7/’:

* CHOCOPHLAN (taxon stratified DNA sequence for the entire biological sequence space): ~/collab/YukaMoroishi/ResultsFiles/HumannDB/chocophlan
* UNIREF90 (protein alignment database): ~/collab/YukaMoroishi/ResultsFiles/HumannDB/uniref
* utility\_mapping (annotations database): ~/collab/YukaMoroishi/ResultsFiles/HumannDB/utility\_mapping
* SILVA: ~/collab/YukaMoroishi/ResultsFiles/HumannDB/
* Human genome: ~/collab/YukaMoroishi/ResultsFiles/HumannDB/

Update databases as needed. Code:

$ humann\_databases --download chocophlan full /path/to/databases --update-config yes

$ humann\_databases --download uniref uniref90\_diamond /path/to/databases --update-config yes

$ humann\_databases --download utility\_mapping full /path/to/databases --update-config yes

**Running HUMAnN3:**

After kneaddata has been run on raw .fastq files, we can run Metaphlan3 and, subsequently, HUMAnN3. Kneaddata files are stored in:

*Insert directory*

Log in to discovery7.dartmouth.edu. If using conda, activate conda environment.

Run Metaphlan and HUMAnN3

*Insert code and directories here*.

Notes:

* Configurations are 1 node, 40 cores, and walltime of \_\_\_, using ‘bigmem’. Change the number of nodes and cores if necessary.
* HUMAnN3 is run on /dartfs-hpc/scratch, and output files and log file are moved to output directory. Temp files are set to be removed after successful run. Scratch files are automatically removed after 45 days.

Output files are saved to:

*Insert directory*

**Running HUMAnN3 on Test Set and Shell Scripts**

This test runs metaphlan3 as well. **Metaphlan3 output is not saved**.

Test humann3 runs using old kneaddata files in:

/dartfs-hpc /rc/lab/M/MRKepistor7/nhbcs/LabResults/Intermediate/kmerged/

/dartfs-hpc/rc/lab/H/HoenA/Lab/Share/MBLmetagenomics/kmerged/

Log in to discovery7.dartmouth.edu. Set directory to:

/dartfs-hpc/rc/lab/M/MRKepistor7/collab/YukaMoroishi/ResultsFiles/bin/

Two shell scripts exist in this folder:

‘run\_sample\_jobs.sh’: Creates jobs for each kneaddata file. Currently set at 100 jobs.

‘test.sh’: For each \_kneaddata.fastq file, runs HUMAnN3 in scratch space. Output files of interest are moved to output directory.

Run following: ./run\_sample\_jobs.sh.

Notes

* 1 node, 8 cores ran out of RAM. Increased number of
* 1 node, 40 cores, ‘bigmem’, 18 hours of walltime 🡪 This runs on N cells
* At most 3 jobs run simultaneously on N cells
* 3-4 hours to run HUMAnN3 for most files
* So far, 6 HUMAnN3 runs per day.
* DO NOT run metaphlan separately while running HUMAnN3. Possible that new chocophlan database will be downloaded when running Metaphlan3. Make sure that ‘chocophlan.v296\_201901 is downloaded. mpa\_v30 or mpa\_lastest do not work so far.